23

OAR 67-012
21 SEPTEMBER 1967

SELECTIVE DISSEMINATION OF INFORMATION *in practice*

Survey of Operational and Experimental SDI Systems

A.G. HOSHOVSKY and C.S. DOWNIE

PUBLISHED BY
OFFICE of AEROSPACE RESEARCH
UNITED STATES AIR FORCE

19

SELECTIVE DISSEMINATION OF INFORMATION IN PRACTICE

Survey of Operational and Experimental SDI Systems

A. G. Hoshovsky & C. S. Downie
Office of Aerospace Research
United States Air Force

Paper presented at the 1967 Congress of the International Federation on Documentation (FID), Tokyo, Japan, 21 September 1967. The views and opinions are those of the authors and do not necessarily represent the position of the Office of Aerospace Research or U. S. Air Force.

INTRODUCTION

Amidst the proposals of the late fifties (1958) a novel and exciting idea was advanced by Hans Peter Luhn of IBM. The idea he proposed, among others, was to exploit the capabilities of computers to provide the users of literature with tailor-made announcements about recent articles in their individual spheres of interest (1).

In the system envisioned by Luhn the computer would compare lists of technical expressions describing user's interests with the technical terms in the abstracts of scientific articles. As matches between words would occur, computers would prepare lists of names of the people who were likely to be interested in the selected documents. The users would be notified about such articles and asked to indicate on a response card the accuracy of the selection. The resultant information would be fed back to the system to adjust the user "profile" to improve the accuracy of the future computer selections.

It is common knowledge that Luhn's suggestion has been tried and implemented in many places. The questions that need to be asked is: How? To what extent? With what degree of success? Cost vs benefit?

The *documentation* literature in which the SDI operators report on their systems provides part of the answer. Another part is now available from two surveys. One survey involved 17 SDI systems. The study served as the background for the design of SDI services in technical libraries of the U. S. Army. It was completed in 1966 (2). More recently another inquiry was initiated by the

Office of Aerospace Research of the U. S. Air Force, to learn more about the various SDI input-output options, with particular emphasis on the economy of operation. This later study is still in progress. To date 39 systems have responded. Of these responses 25 have been analyzed. The facts which have emerged from this inquiry, coupled with the findings of others, are presented in this paper.

GENERAL OVERVIEW

We would like to start the general overview of these systems by a few remarks about SDI as it should be viewed in the context of information-acquiring habits of scientists and engineers. Imagine, if you will, the bull's eye of a target. The area in the center of the bull's eye represents the information the scientist regularly receives from his associates and the journals he subscribes to. As the circles widen out, the probability of coming across relevant information becomes increasingly smaller. Items found in obscure journals, foreign language publications, or reports of smaller industrial R&D establishments usually fall well outside of the bull's eye and thus escape the user's immediate attention. This is where SDI offers its greatest help - by increasing the user's peripheral vision, as it were.

With these preliminary remarks out of the way we want to move to some observations about the present SDI systems.

- While it may sound strange, SDI has been, and still is very much the American information technique. As far as we know (and we could be mistaken) the only countries that have invested in SDI

based systems are the U.S. and U.K. Our search through literature and interviews has failed to uncover a single operating SDI system in any other country.

- If the various sources of literature are to be believed, we should have more than 70 systems in operation. However, we have identified and verified only 39 systems. There are still several to be examined; however, we do not expect the total to be much above 45. These 45 systems serve an approximate population of 30,000 users. Many of these users are served through group profiles, hence the actual number of profiles that are being processed through the SDI systems is somewhat lower. Together the SDI systems process some 100,000 individual items of literature each month.

- As many as 58% of the systems are viewed by their designers or operators as fully operational; 13% as partially operational. Only 29% are considered experimental, i.e., in a stage of development where their feasibility and utility is still under serious questioning.

- The average age of a fully operational system is 3.5 years; the oldest of these is now 6.3 years old. The average age of the experimental system is less than one year.

- Significantly, practically all of the systems have been established by such institutions as government agencies, large industrial organizations and large information services. Our sample revealed

only one SDI system set up by a university at the initiative of its library.

- Almost one-half of the systems (42%) depend totally on the preparation of their inputs by the major information systems of the National Aeronautics and Space Administration, Atomic Energy Commission, National Library of Medicine, Chemical Abstracts and the Institute of Scientific Information.

- 92% of the systems use computers to perform the profile-document matching operations. The equipment used so far ranges all the way from the IBM 1401 to the IBM 360.

- Only one system in the sample of 25 has been designed solely for technical managers; four (16%) are used exclusively by scientists. The remaining 80% are used to provide services to both scientists and technical managers

- There are wide differences between the systems with respect to number of document entries they process. The largest system claims to process as many as 30,000 entries each month, while the smallest runs no more than 150. Within these boundaries it appears that there are two major groupings. About 60% of the systems process anywhere between 150 - 1,200 documents per month - while the other 40% process anywhere between 1,200 and 300,000. As might be expected the large processors are AEC, NASA, and the Department of Defense.

- The present systems operate on profiles of single individuals as well as groups. Forty (40%) of the systems serve individuals exclusively. Another 40% processes both the individual and group profiles. Only 20% are designed to serve the needs of groups alone.

- Looking at the character of the groups served by the SDI systems we find that they fall into three categories: administrative groupings such as whole companies; the task/mission oriented groups which cover a wide range of disciplines; and the science oriented groups whose interest usually centers in narrower and more easily identifiable segments of literature. The breakout within our sample is as follows: by administrative grouping - 21%, by task orientation - 57%, by discipline - 57%. The reason these figures add to more than 100% is because some of the systems provide services to more than one group.

- Only about 25% of the systems use profiles developed through the free selection of terms without reference to a thesaurus. The others require selection of terms from a controlled vocabulary.

SPECIAL OBSERVATIONS

- The Costs are Appreciable but Not Unreasonable

  Reliable costs of SDI services are difficult to obtain, partly because the usual SDI service is a part of the larger services provided by a system and partly because of the lack of standard definition. However, there are some indicators that the

approximate average cost of serving a single user is about $100 per year. For example the estimated cost of operating a service using the SDI-3 program on an IBM 1401 system, processing 1,200 documents against 500 profiles is $92.50 per profile. (3:31) The cost of the selective dissemination system of the Institute of Scientific Information can be obtained at a flat basic rate of $100.00 per user per year. Similarly, the cost of providing service to some 200 Air Force users enrolled in the NASA SDI system amounted in 1966 to approximately $80.00 per user.

- Admittedly these are relatively high costs, but not unreasonable, if one considers that the SDI service is capable of saving about 1% of the technical man's time. (3:33) Assuming an average cost of employing a technical man to be $15,000 per year -- this represents a saving of $150.00; in our opinion a fair return on an investment of $80 to $100. The use of group profiles would, of course, make these savings more impressive. But the really major savings are not represented by the time saved. Rather, they are in providing the technical man with what we have earlier called peripheral vision, i.e., the ability of learning about information which could be essential to his work, but which is published in the literature he normally would not consider as likely to contain anything of interest. In this connection there is some evidence that certain users see very little of this peripheral information. The measurement of engineers at one installation exposed to an SDI

system showed that only 10% of the material supplied by the SDI system had been seen before (8).

- User Satisfaction is High

The measure of user satisfaction with SDI is an elusive concept - measured mostly on subjective terms. Our own earlier study in 1966 indicates, however, that it is unusually high. Approximately 96% of our respondents have indicated that they consider the service either useful or highly useful. They back up this opinion with an indication of their willingness to allocate a portion of their budget to continue the SDI Services. (3:33)

- Satisfaction Correlates Positively with the Volume of Notifications - Rather than with the Precision of Matching Profiles

Contrary to the usual belief, no correlation has been determined between the degree of satisfaction (expressed in terms of time saved and expansion of coverage) and the accuracy of the matching. In one study users with the matching precision percentage as low as 25% wanted to continue the service because it expanded their literature searching ability and saved them time. Conversely, users who were not completely satisfied with the system were receiving notices with the matching precision of 47% and higher. (4) Thus user satisfaction with the degree of precision in matching his individual interest profile depends mainly on his personal search preference -- whether he will accept less precision in

matching his profile in return for greater coverage. We suspect that this is largely a question of trade-off; how much of his own time is the individual user willing to invest in final screening of the documents in order to broaden his coverage? For example, starting with a thousand documents to be screened by SDI techniques, three different users may indicate three rather divergent preferences.

	Number of Documents (of the thousand)			
User	Received	Relevant	Non-Relevant	Screening Precision %
A	3	3	0	100%
B	30	15	15	50%
C	100	30	70	30%

In the same study the author shows that there is a positive correlation between the degree of satisfaction and the literature coverage provided by the SDI service.

Similarly, in a study of 45 IBM research scientists Resnick and Hensley (5:109) found that users do not want to place a limit on the number of notifications received. 30% of them indicated that the number of notifications received during the experiment was too high, but only 15% of subjects indicated that this was one of the features they did not like.

- SDI Applications

By far the largest area of application is in providing current awareness to scientists and engineers. There are other applications, however, some of which offer immediate and tangible benefits to the organization.

One of the earliest such applications was made by IBM to provide information service to marketing personnel rather than to research and development people. Kraft describes such a system used by the IBM Midwestern Region based in Chicago. (6) The input publications consist of such items as the Wall Street Journal, Fortune, Sales Management, Control Engineering, etc. In another case an aerospace company is looking into the possibility of using SDI to disseminate company reports to technical managers in government. Presumably greater knowledge on the part of government personnel about the competence and products of a given company might improve the chances for successful competition on government contracts.

A novel use of SDI, and in our opinion one of far-reaching significance, is its recent application to the control of library acquisitions. Mr. Carlos Segarra, of the Army Engineer Research and Development Laboratories, is using profiles of laboratory projects, instead of people, to reflect the information needed to support the particular project or task (7). The resultant

profile for the Center consists of approximately 3,000 terms, updated at 6 months' intervals. The profiles act as a filter for the incoming documentation, rejecting those documents from further processing into the library's permanent collection which could not be matched up with the Center's interest. Segarra claims to have achieved savings of $3.60 per processed document (reduction from $7.66 to $4.06 per report). Additionally this technique reduced processing time from 75 days to 5 days and significantly increased the use of holdings.

OBSERVABLE TRENDS

Conclusive data is generally lacking to establish undisputable trends. Even so, certain actions on the part of the SDI operators suggest forces at work in several directions.

- Efforts to Capitalize on the Products of Large Systems

The high costs of identifying, organizing, indexing, and keyboarding information about the documents, coupled with the evergrowing demand for coverage of a wide subject area are the obvious motives for the efforts of maximum utilization of tapes and indexes by large systems. SDI systems which take inputs from several such systems are now in evidence, and are expected to grow.

- Increased Use of Group Profiles

Another significant development is the trend away from individual

profiles to group and standard profiles. The evidence of this trend is present in the recent NASA emphasis on their new SCAN system, which provides information on some 189 areas, covering the 34 subject categories used in STAR (Scientific and Technical Aerospace Reports) and IAA (International Aerospace Abstracts). It is also shown by the customer preference at the Aerospace Research Application Center (ARAC), a self-supporting private SDI service which disseminates government-generated information to industry. When given the option to subscribe either to individual profiles at a cost of some $200 for up to twenty notifications, or to standard profiles at $80 per year, 70% of them elected standard group profiles. The trend is being further established by the recent experiments conducted jointly by the Air Force Office of Aerospace Research and the Clearinghouse for Federal Scientific and Technical Information. The basic method used in this experiment is to break out the Clearinghouse announcement bulletin into some 65 categories to permit receipt of only those announcements which fall within the users area of interest. This experimental announcement program known as CAST (Clearinghouse Announcement in Science and Technology) began in September 1967. 950 Air Force scientists and engineers were included. If proven successful it may be opened to general subscription.

- The Trend to Commercial SDI Subscription Services

The trend, which we believe is of greatest significance, is toward the emergence of private services willing to provide selective announcements to anyone who is prepared to pay the service fee. Perhaps the best known is the ASCA (Automatic Subject Citation Alert), an SDI service of the Institute for Scientific Information (ISI) in Philadelphia. ISI has been testing and operating the SDI system for the past three years. The test involved more than 500 scientists. The system covers annually about 300,000 articles in some 1,600 journals. The basic annual subscription costs $100.00.

Numerous other centers, especially those established under the auspices of NASA for the purpose of promoting transfer of space age advances to the private industrial sector, provide SDI services to the companies and to private individuals. ARAC, mentioned earlier, is one of such centers which provides a commercial service at a cost of $80/per standard profile. Recent efforts by the Chemical Abstracts Service to provide their tapes for SDI exploitation by major companies lends further indication of a trend which is most encouraging.

CONCLUSIONS

- Considering the fact that there has been a steady and rapid growth of SDI systems; that many of these systems have been in existence for a number of years; that there has been a corresponding growth

of SDI users; and that SDI is now made available through subscription from commercial sources, SDI appears to be becoming accepted as an effective and economical method of assisting scientists and engineers in keeping abreast of advances in their fields.

- As evidenced by the respondents, who estimate that 90% of the material they receive from SDI was information they had not previously seen, SDI is especially valuable in providing the user with peripheral vision of information which is of direct interest to him.
- The relatively large computer and other costs of installing and operating SDI systems tend to inhibit any but the larger organizations from establishing SDI services. The availability of free computer time for initial experimentation is usually an essential springboard for the start of the service.
- SDI systems appear to be developing in two major directions: (1) greater use of machinable records, produced by large documentation systems, and (2) experimentation with group profiles. Both trends promise significant reductions of operating costs.
- An encouraging trend also is found in the availability of SDI services through the commercial subscriptions from such organizations as Chemical Abstracts or ISI. This promises to bring SDI within the range of the individual or small organization anywhere in the world.

- Both the availability of computer tapes and direct SDI services should permit greater use of SDI by countries who may not as yet have resources or facilities to develop and build independent services of their own.

REFERENCES

1. Luhn, H. P. "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment." AM. Document. 12 (2) 131-38 (1961)

2. Binova, W. A. and E. J. Golblum. Selective Dissemination of Information: Review of Selected Systems and a Design for Army Technical Libraries, Information Dynamics Corporation August 1966 (Clearinghouse, AD 636 916)

3. Downie, C. S. et al. The Office of Aerospace Research Scientific and Technical Information Program, U. S. Air Force, 15 March 1967 (Clearinghouse, AD 656 694)

4. Sage, C. R. "Computer Dissemination of Current Literature - An Overall Evaluation," Paper presented at Annual Meeting of American Society for Engineering Education, 20-24 June 1966, Washington State University, Pullman, Washington.

5. Resnick, A. and C. B. Hensley. "The Use of Diary and Interview Techniques in Evaluating a System for Disseminating Technical Information," AM. Document. 14 (2), 109-16 (1963)

6. Kraft, D. H. "An Operational Selective Dissemination of Information: (SDI) System for Technical and Non-Technical Personnel Using Automatic Indexing Techniques." Proc Am. Document. Institute, Chicago, Ill.. Oct 6-11, 1963, 69-60.

7. Segarra, C. O., An Approach to Cost Effectiveness of a Selective Mechanized Document Processing System, ERDL, U. S. Army, March 1967 (Clearinghouse, AD 651 486)

8. Bromberg, E. I., "The Bonneville Power Administration Selective Dissemination of Information Program." A paper presented at the IEEE Summer Power Meeting, Portland, Oregon, July 9-14, 1967.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Office of Aerospace Research
Information Studies Division
Office of Scientific and Technical Information
Arlington, Virginia 22209

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
Selective Dissemination of Information in Practice:
Survey of Operational & Experimental SDI Systems.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Management State-of-the-art Special

5. AUTHOR(S) (First name, middle initial, last name)
Alexander G. Hoshovsky & Currie S. Downie, Colonel, USAF

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
21 September 1967	15	8

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	OAR 67-0012

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Tech-other	Office of Aerospace Research 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

The primary purpose of the report is to present an overview of the operational and experimental systems established for the selective dissemination of scientific and technical information. Secondarily an attempt has been made to identify the trends which may shape the future development of the selective dissemination procedures. The report is based in part on the existing SDI literature and in part on the results of two recent surveys. The combined results indicate that there are approximately 45 SDI systems in various stages of operation, serving an approximate population of 30,000 users. The largest system processes as many as 30,000 entries per month, while the smallest runs no more than 150. AEC, NASA and DOD are the three largest processors. Operating costs are in the vicinity of $100 per user per year. The user satisfaction is high. This satisfaction correlates postively with the degree of literature coverage. There seems to be no correlation with the precision of matching. The observable trends are toward the efforts of capitalizing on products from large systems, toward an increased use of group profiles and toward commercial subscription services.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
SELECTIVE DISSEMINATION OF INFORMATION (SDI)						
DISSEMINATION OF INFORMATION						
SCIENTIFIC INFORMATION						
COMPUTER SYSTEMS						
INFORMATION STORAGE AND RETRIEVAL						
LIBRARIES						
DOCUMENTATION						
MECHANIZATION OF SCIENTIFIC DOCUMENTATION						
INFORMATION AND COMPUTER SCIENCES						